1974

Differences Between Supervisor and High and Low-Rated Employees' Perceptions of Job Performance Ratings and Importance of Job Factors

Harvey Edward McGowan
Portland State University

AN ABSTRACT OF THE THESIS OF Harvey Edward McGowan for the Master of Science in Psychology presented February 8, 1974.

Title: Differences between Supervisor and High and Low-rated Employees' Perceptions of Job Performance Ratings and Importance of Job Factors.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

[redacted]

Milton K. Davis, Chairman

[redacted]

Robert F. Powloski

[redacted]

James A. Paulson

This study was an attempt to gain insight into differing conceptions of job performance and job performance factors held by supervisors, employees in general, and of employees rated high and low on overall job performance by their supervisors. The discrepancy in the perceptions of job performance is an element in a general pattern of a well-documented discrepancy in the perception of subordinates' wants, needs and desires by superiors.

To assess employees' perceptions about their own job performance, self-ratings were taken, along with estimate ratings of how employees thought their supervisors would rate them. A graphic rating scale was used, Format III, with seven job performance factors; Ability to work with others, Amount of work done, Quality of work done, Leadership potential, Ability to do complicated jobs, Ability to work with minimum supervision, Conscientiousness, and an eighth scale, Overall Performance. Subjects were 78 female assembly workers along with their eight immediate male supervisors.

It was hypothesized, on the basis of earlier studies, that employees would rate themselves higher than their estimate ratings, which , in turn, would be higher than actaul supervisor ratings. This general ordering was found to be significant for all factors except 1,2 and 3 at the .05 level and beyond. The ordering of scores on scale 8, Overall performance, was significant at the .01 level.

Differences were explored among groups of employees in the degree of discrepancy between estimate ratings and supervisors' ratings. Employees were divided into high and low groups according to ratings given them by their supervisors; both groups consisting of one-third the total number of employees. It was hypothesized that low-rated employees would show more discrepancy than would high-rated employees, and both groups would overestimate their

overall performance.

Though the discrepancies for the two groups were in opposite directions (higher group underestimating, lower group overestimating), the difference in the absolute amount of discrepancy did not approach significance. The lower group did not show significantly more discrepancy as predicted. Also, the high group rated themselves and estimated their ratings significantly lower than actual supervisors' ratings.

The third hypothesis was based on the assumption that high-rated employees are so rated partly because the relative importance these employees attach to job factors is more similar to that attached by supervisors than are those by low-rated employees. Multiple regression equations were computed for supervisors' ratings, high-self ratings and low-self ratings using the overall performance rating as the criterion and the other seven factors as predictors. As predicted, high-rated employees demonstrated closer correspondence in their weights to those given by supervisors than did low-rated employees.

Conclusions based on the study, limited by nature and number of subjects and the type and number of factors used in the rating scale were as follows; a) employees, as a whole, rate themselves higher both on self-ratings and on estimates of their supervisors' ratings than the supervisors actually do rate them, b) neither high nor low-rated employees have any better idea

of how their supervisors rate them on overall performance; both groups show approximately the same amount of discrepancy between their estimate of their supervisor's ratings and the actual ratings with high-rated employees underestimating themselves and low-rated employees overestimating themselves on overall performance, d) high-rated employees seem to have a better idea of the job factors considered important in their contribution to overall performance by supervisors than do low-rated employees; the factors high-rated employees judge as important are generally the same factors their supervisors judge as important with a few notable exceptions.

DIFFERENCES BETWEEN SUPERVISOR AND HIGH AND LOW-RATED EMPLOYEES' PERCEPTIONS OF JOB PERFORMANCE RATINGS AND IMPORTANCE OF JOB FACTORS

by

HARVEY EDWARD MCGOWAN

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
PSYCHOLOGY

Portland State University
1973

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the thesis of Harvey Edward McGowan presented November 13, 1973.

Milton K. Davis, Chairman

Robert F. Powloski

James A. Paulson

APPROVED:

Ronald E. Smith, Head, Department of Psychology

David T. Clark, Dean of Graduate Studies and Research

ACKNOWLEDGEMENTS

I am indebted to a number of individuals without whose time and efforts, the completion of this project would have been questionable; to Dr. Milton Davis, my thesis advisor and committee chairman, who has lent his wisdom and guidance to the development of this study from its initial stages; to Drs. Robert Powloski and James Paulson, whose helpful suggestions and criticisms aided in the formulation of methodology and analysis; to Mr. Ed Srebnik and Jack Morehouse of Tektronix, Inc. who so willingly and enthusiastically demonstrated their support for personnel research by arranging the availability of employees and the facilities for collection of data; to those women at Tektronix who contributed their time, efforts and, in some cases, their coffee breaks, so that some stranger from Portland State University might conduct a bit of research; and, of course, to my wife Kathy, whose support, encouragement and understanding have never failed me.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The assessment of human behavior is a problem not exclusively confined to the psychologist or to the realm of psychology. It is quite apparent that psychologists are not the only people involved in the formal measurement or evaluation of human behavior. Educators, for example, are asked to periodically assess the behavior of their students in a systematic, written manner. These assessments are, of course, subject to the same imperfections and failings of all measurements of human behavior, but nonetheless serve an immediate purpose and affect the lives of a good many individuals. Those engaged in the management and supervision of human behavior in an industrial setting also attempt to formally assess the behavior of individuals. This latter form of assessment is, in part, the subject of this study; a form of assessment which not only possesses all the shortcomings of measurement techniques in general, but also has inherent imperfections and yields an uncanny influence over the lives of millions of working men and women.

As Tiffin and McCormick (1965) point out, the rating of working men by their supervisors is by no means a recent development. Supervisors have always rated their workers,

and these ratings, however slipshod and unsystematic, have always had enormous influence on personnel decisions. Fortunately, they have become increasingly more sophisticated, resulting, among other things, in more comparability from one employee to another. Formal rating systems have not ceased, however, to be a source of continuing controversy. It is a basic tenet of human nature and psychology that human beings do not perceive all things similarly. It would be expected, then, that a ratee would have a very different perception of his job performance than would a rater, by the very fact they are different human beings with different values, expectations and viewpoints. The nature and degree of this discrepancy between rater and ratee and the consideration of group differences among ratees is the subject of this study.

CHAPTER II

PERFORMANCE APPRAISAL TECHNIQUES

I. USES OF PERFORMANCE APPRAISAL

Psychologists develop measures of human behavior primarily to understand more about the behavior being assessed or measured. When these same measuring techniques are used in an applied situation, they are often implemental in making practical decisions about human beings. This is certainly the case with merit rating systems. Understandably, analyzing the behavior of employees is secondary to obtaining a measure of their performance and making decisions based on that measure. This fact becomes more apparent when one looks at the frequency with which merit rating systems are used for administrative purposes as compared with research purposes. The National Industrial Board (1954) found that, of those companies surveyed which were using some sort of rating system, 72% used them in the administration of wages, 63% for identifying promotable employees, while only 22% used them for any kind of research. Barrett (1966) lists three general purposes for which rating systems find use in industry: administrative--programs directed at some immediate personnel action such as salary determination,

promotion or layoff; performance improvement--primarily through supervisory review and research; and research--particularly to ascertain the value of selection and training procedures. As might be expected, the primary general purpose of rating systems is administrative, according to Barrett, with salary administration being the single most important use. Benjamin (1952) also found the single most common use of ratings to be in connection with merit pay increases.

Although rating systems are used for other purposes such as employee self-improvement, to determine training needs among employees, and to sort special talents among employees, the fact remains that merit rating systems are, by their very nature, evaluative. Merit rating systems attempt to directly determine an individual's value or usefulness in a particular organizational setting. While an intelligence test might demonstrate to an individual that he does not possess some skills which society tends to value, it does not make direct, evaluative statements about his worth as a functioning human being in a setting that is tied to his very livelihood. While rating systems are used for a variety of reasons which may differ from company to company, the primary purpose of these systems is the determination of a particular employee's worth to the company, a determination upon which is based the administration of wages and the judgment of promotability.

II. TYPES OF PERFORMANCE APPRAISAL TECHNIQUES

Just how is the effectiveness of employees currently measured in industry ? A brief description of various methods of performance appraisal or merit rating systems follows:

Graphic Rating Scales

The most widely used method in which a series of trait names, factors or characteristics are listed and a mark of some kind is placed on a line or at intervals which represent the degree to which the item is exhibited or possessed by the individual being rated.

Rank Order Method

With this method, all employees rated by the supervisor are ranked from best to poorest in one or more traits.

Forced Distribution Method

Individuals are distributed along one or more scales and a fixed percentage is assigned to the best and worst ends of the distribution as well as to the middle bracket of the distribution.

Paired Comparison Method

For each trait every subordinate is compared with every other subordinate and the number of favorable choices is tallied for each individual.

Critical Incident Method

This system involves the keeping of a record of unusually good or undesirable incidents occuring in an employee's work.

Forced Choice Method

Sets of descriptive statements are given and the rater selects those which are most and least descriptive of the employee.

Free Written Essay

Someone in a position to know the ratee's work or performance is asked to describe that performance and evaluate it.

Objective Measures

The most easily measured aspects of performance such as sales, amount of pieces produced, or scrap rate are used as a basis of evaluation.

Weighted Random Checklists

A series of brief descriptive phrases are arranged that have been scaled in favorableness and assigned scoring weights not known to the rater who checks items which apply to the individual being rated, leaving others blank. The total value of those items checked plus a correction factor gives the final rating score.

Each of these methods has advantages and disadvantages of their own. Generally, it can be said that these methods, by one means or another, attempt to compare individuals with one another or with some objective scale. The most widely used method, the graphic rating scale, has certain advantages in that it allows comparison not only with a presumably objective scale on any number of traits or characteristics, but also allows comparisons among individuals rated on the same sclae. No other rating system allows this. For these reasons and others, this method of evaluation was chosen for use in this study.

III. EVALUATION OF APPRAISAL SYSTEMS

The basic question one must ask about any measuring device, whether it be a yardstick, an aptitude test, or a merit rating system, is "Does it measure accurately?". This question becomes increasingly more difficult to answer concerning any device which purports to measure facets of human behavior. Fortunately, in the process of developing techniques of measuring behavior, behavioral scientists have developed fairly sophisticated methods for determining an instrument's value; how stable the instrument is, how susceptible it is to random variables, and how well it corresponds to other instruments or, in a practical situation, to certain desirable criteria.

The evaluation of merit rating systems in industry

poses special problems. For one, merit rating systems usually are not originated by people concerned with scientific methods, but rather by a pragmatic industrial management who desire a fast, somewhat permanent evaluation of human beings as units which expend certain amounts of energy in contributing organizational goals. If a merit rating system seems to serve these purposes and does not seem to be grossly inaccurate, then there is no reason to change the system. Thus, rating systems are not generally submitted to the same rigorous tests that other devices are, and so are denied a process of development and refinement. Another problem is that there exist few criteria outside the ratings themselves with which merit ratings can be compared for purposes of validation. Merit ratings themselves are used as criteria in a great many studies in personnel and industrial psychology.

IV. THE GRAPHIC RATING SCALE

As stated above, a graphic rating scale was chosen for use in this study because of certain advantages it possesses over other rating systems. It allows us to examine a quantitative measure of the degree to which an individual is attributed with a certain trait or factor. Furthermore, since individuals are rated using the same scale, they can be compared to one another. Scores on each of the several traits or characteristics used can be compared to general,

overall measure of employee performance. Conceivably, the contribution of certain factors to general job performance could be assessed and group differences be explored in this respect. Not the least of the advantages of using this particular form of appraisal is the fact that it probably possesses the largest backlog of related research of any rating system, and, as a result, has gone through a process of revision and refinement..

Problems

There are a few basic problems inherent in the graphic rating scale which a good number of researchers have investigated. There is the familiar "halo" effect of which Thorndike (1920) remarked, "ratings are apparently affected by a marked tendency to think of a person in general as rather good or inferior and to color judgments by this general feeling." If the halo effect is at work in the ratings, there results a higher correlation among traits than would be expected otherwise. Generally, it is held that this effect tends to detract from the accuracy of the rating scale. However, some investigators, including Bingham (1939), believe that the effect is not really as notorious as is commonly held. He states that a similar effect is present in the perceptual process and in the act of judgment about any human being. It is unreasonable, he asserts, to expect raters to give a pure evaluation of each factor by itself, since all traits and characteristics of an individual

combine in determining how an individual's personality is seen by others. Nevertheless, one method has often been suggested (Barrett,1966 and Tiffin and McCormick,1958) to eliminate or at least subdue the halo effect. This method involves having the raters rate each individual on a particular trait before moving on to the next trait. This system supposedly forces the rater to think of one trait as possessed by each individual, one at a time, instead of thiking of the whole person. Taylor and Hastman (1956) compared this method with the traditional method of rating an individual on all traits before going on to the next individual. They found the results of the two methods to be essentiaaly the same and suggested that the traditional method be used over the more involved and time consuming alternative method.

Another inherent problem of graphic scales is commonly referred to as the error of central tendency. This is the tendency of raters to use the center of the scale and to avoid extremes. Although that being rated is generally assumed to be normally distributed, involving a majority of the cases clustered around the mean of the distribution, the effect of central tendency is to assign more cases to the middle section than is actually warranted. It is difficult, of course, to determine which ratings actually belong there from those placed there for another reason. Generally, it is held that, as Barrett (1966) states, the

error of central tendency occurs when the rater is unsure about what is being rated or does not have sufficient knowledge about the ratee's behavior to make any valid judgments about it. The unwritten rule seems to be, "when in doubt, use the middle.". So the solution to this problem of central tendency would seem to be to remove the doubt. This can be accomplished to some degree by first eliminating from consideration those with whose behavior the ratee is unfamiliar, and by making it clear what type of behavior is representative of the points on the scale. It is hoped that the error of central tendency can be minimized by the use of such methods.

There exists another tendency on the part of raters to rate the majority of ratees as being above average. This tendency, referred to as the leniency effect, is oftentimes the result of personality factors within the individual doing the rating rather than a result of something within the rating system itself. This tendency becomes more pronounced when it is known that the ratee will see the ratings or when it is known that the ratings will directly or indirectly influence the future of the ratee. Although leniency can probably never be eliminated completely (educators have forever been plagued with the "C" student being inevitably below average), there are measures that can be taken to assure that this tendency can at least be minimized. This can be accomplished quite easily when the ratings are to be

used in research by assuring that the ratings will be kept confidential and will not be used in making decisions about those being rated. It is hoped that this will enable the raters to be more candid and truthful in their ratings than in the typical circumstances surrounding merit ratings.

Research and Desirable Qualities

Thus far, we have seen some of the problems associated with the graphic rating scale and how these problems might be minimized, if not completely eliminated. These have been problems connected with almost any kind of graphic rating scale. There are, of course, some scales or types of scales constructed in such a manner that they minimize any problems and also retain that quality which every measure of human behavior must possess; namely, reliability. According to the old rule of thumb, if a measuring device is not first reliable, it can never be valid. As was discussed above, validity is a special problem for merit rating systems and is beyond the scope of this study. And so, as a result, we should be concerned with reliability and should choose a type of rating scale which has been shown to possess a satisfactory amount of reliability. We should also explore other qualities that a rating scale should possess which eventually would be considered in the final choice of a rating scale for use in this study.

Graphic rating scales can differ from one another on a number of dimensions or qualities; some of which are the

number of traits or factors to be rated, number of rating scale categories, and type of scale anchoring. Any graphic rating scale used in research or in a practical situation should embody those characteristics which have been shown to be superior.

Empirical attempts to deterine the number of factors related to job performance have been somewhat inconclusive. Ewart, Seashore, and Tiffin (1941) performed a factor analysis on the results of a 12 trait merit rating scale used to rate 1120 men and found only two factors; ability to do the present job and the possession of skills above the requirements of the job. A third factor, health, was regarded by the investigators as an "artifact". Roach (1956) had managers describe supervisors using a 390 item checklist questionnaire and, using factor analysis, found 15 factors. Grant (1955) asked supervisors to rate managers on an assignment-type rating scale with 20 assignments or job requirements listed and found one "general" factor and five "group" factors. These examples are enough to illustrate briefly the lack of agreement concerning the true number of factors involved. Seashore, Indik and Georgopoulos (1960) conclude that there is little support for the notion that there may exist some generalizable pattern or set of patterns describing the composition of job performance and the relationship among components of job performance. They point out that the relationships among different aspects of

job performance are generally small and that the size and direction of these relationships are, to a large degree, unique to each population and situation, and somewhat different for organizations as opposed to individuals.

The question arises as to whether one should, in choosing the number of factors to be rated, pay heed to the confusing state of research on job performance factors. It is doubtful whether an employee's job performance can be thoroughly assessed using as few as two factors. On the other hand, ratings requested on as many as 15 factors might well create a difficult and confusing task for raters with questionable results. Indeed, the majority of rating scales contain between five and eight factors. The advice given by Bittner (1950) seems sensible and appropriate. He states that traits or factors should be selected on the basis of the following; observability-can the rater actually observe the trait?, universality-is it an importnat characteristic of successful performance on all jobs to be rated?, and distinguishability-is the trait clearly distinguishable as meaning something different from another trait with another name?. These three general criteria have been considered in the present study in the eventual choice of a rating scale.

In an early article, Symonds (1924) used a statistical argument to assert that, in regard to the number of rating categories to be used in a rating scale, reliability

should increase somewhat with the number of categories, but negligibly so when more than nine categories are used. Bendig, however, in two studies (1953,1954a) varied the number of rating scale categories and compared rater reliability. In the first study, he compared scales using 3,5,7,9 and 11 categories. He found essentially equal reliabilities for all numbers of categories with the exception of the scale using eleven categories, in which case reliability decreased somewhat. In the second study, he compared 2,3,5 and 7 categories and again found no consistent relationship between number of categories and reliability. He maintains that Symond's argument may hold true for test reliability, while it is rater reliability which is the concern of most people who investigate rating scale methods. Apparently, it can be said that the number of rating scale categories has little influence on rater reliability.

In his first study, Bendig also varied the amount of verbal anchoring, or verbal descriptions of the points on a rating scale. Three conditions were compared; the center category was defined, both end categories were defined, or both end and center categories were defined. He found that the reliability of the scale increased with the amount of verbal anchoring. Peters and McCormick (1966) compared the reliability of numerically anchored scales with that of verbally anchored scales. Ratings were made on five sensory/physical dimensions of job activities. Reliabilities

for all scales were computed by the analysis of variance approach. In a test of statistical significance across all five dimensions, it was found that job-task anchored scales could generally be used with significantly greater reliability than numerically anchored scales. Stockford and Bissell (1949) sought to improve a rating system for Lockheed Aircraft Corporation. The original scale used an evaluational scheme of anchoring. The investigators replaced the original scale with a descriptive scale using behavioral anchorings. They concluded that ratings on descriptive scales were more reliable, less influenced by bias, and show less deviation between raters in leniency and severity than is characteristic of ratings done on evaluational scales.

Thus, from the research we have cited so far, it appears that some characteristics of graphic rating scales are definitely superior, while other characteristics demonstrate little or no advantage, the inclusion of which would be dependent upon personal preference of the person constructing the scale or the person doing the actual rating. It can be said, then, that as far as the number of traits or factors is concerned, that a relatively small number of traits or factors fails to elicit a thorough assessment of an employee's job performance, while a large number of factors serve to confuse the rater and do not substantially facilitate the rating process, As a result, the majority of

rating scales employ a moderate number of factors, usually from five to eight. Number of rating scale categories seems to have little influence on the reliability of rating scales, and would seem to be dependent also on the personal preference of those constructing the scale. Type of scale anchoring, however, does seem to have a great deal of influence on reliability. It is apparent, first of all, that verbal anchoring is superior to numerical anchoring and that the more scale categories anchored verbally, the better. Also, it has been shown that behavioral anchoring is superior to simple evaluative anchoring. For the purpose of this study, it was decided to choose a rating scale which possessed characteristics which have been shown to be superior by research and which had had research performed on it itself. The graphic rating scale eventually chosen for the study was Format III, published and developed by Personnel Research and Development Corporation; Cleveland, Ohio.

CHAPTER III

FORMAT III: RESEARCH

Format III was the subject, along with three alternative formats, of a series of studies performed by Barrett, Taylor, Parker and Martens (1958). The four formats, known as Formats I,II,III and IV, varied from unstructured to highly structured, and were tested for inter-rater reliability, halo, leniency and variability. All four formats attempted to obtain measurements on seven traits or characteristics;

1. Ability to work with others,
2. Amount of work done,
3. Quality of work done,
4. Leadership potential,
5. Ability to do complicated jobs,
6. Ability to work with minimum supervision,
7. Conscientiousness, and
8. Overall performance.

Format I used only trait names with no legend or numerical scale provided. Format II used a verbal description of the several traits instead of trait titles, again with no legend or numerical guide provided. Format III contained trait names, no trait descriptions, and behavioral descriptions of the kind of behavior which characterized each of the five major divisions of the scale. Format IV contained trait descriptions but no trait names and the same behavioral descriptions of the points of the scale used

in Format III. On all four formats, raters placed a check-mark at the desired point on a ten inch line divided by vertical markers into five major divisions, which, in turn, were subdivided into three equal sections with no legend or numerical scales provided. Intrarater reliability coefficients were computed on each of the eight scales for the four different formats. The average correlation for all formats ranged from .51 to .67 with Format III clearly the best. The reliabilities for all scales were uniformly higher for Format III than for the other formats. The investigators wrote;

> The greater stability of Format III makes it more widely useful than the other formats since it might happen that an investigator would be more interested in the information in one of the areas which is unreliably measured by one of the other formats. On Format III he would always have available a measure with adequate reliability for most purposes. (p. 340)

Raters also showed less leniency on Format III when means and measures of skewness of each of the four formats were compared. Halo was measured by taking the average intercorrelations of each of the four formats. Format III was shown to have a slight superiority over Format IV in this respect. Variability was measured by computing the means and standard deviations of each format. It was decided beforehand that with a mean of 8.0, a standard deviation of 2.5 would have been ideal, making it possible to extend the scale from -3 to +3. Instead of following the pattern usually expected of this type of rating scale,

namely, that the variability would be less than desired, it was found that most standard deviations clustered between 2.4 and 3.0. The investigators themselves were a bit surprised at this; "It turned out, however that the supervisors' ratings were, on the average, slightly more variable than had been anticipated even when a trait name had been used." They propose that perhaps the instructions given to the raters may have had some influence on this phenomenon. Raters were told to use all the steps of the scale and not to hesitate to use the extremes if they reflected the true behavior of the ratee. Since all the measures of variability were very similar for all the formats, the investigators concluded that the differences in formats did not effect variability systematically.

Format III was used by these same researchers in a later study to be discussed which was, in part, the basis of the present study. In summary, for the above reason and because Format III seems to possess all the necessary qualities for a suitable research tool, permission was sought and obtained from Personnel Research and Development Corporation (PRADCO) for the use of this format.

CHAPTER IV

THE PROBLEM

I. RATER AND RATEE CONFLICT

Unlike most other measurements of human behavior, performance appraisal systems, as used in industry, are, in varying degrees, series of subjective judgments by someone other than by the ratee. The subject of other measuring techniques, however, is directly involved in the appraisal process. It is the subject's own responses on the MMPI, for example, which determine his personality profile, and it is his responses which determine his score on a general ability test. While testees in these circumstances may disagree with the validity or importance of such measures, he is not entirely subject to another person's appraisal of his performance as he is with most rating systems used in industry. While most employees would agree that some form of appraisal system is necessary, and do appreciate knowing where they stand, it is inconceivable that they would entirely agree with some of the end results of such a system, no matter how it is implemented. Hence, when rating results are communicated, typically within the structure of an appraisal interview, conflicts are bound to arise due to the differing perceptions, attitudes and needs

which rater and ratee bring to the interview. Meyer, Kay and French (1964), reviewing performance appraisal techniques at General Electric Company, found that the traditional interview technique consisted mainly of the rater justifying his ratings and the ratee acting with defensiveness to any criticism on the part of the rater. As would be expected, they found that the more criticism a man received, the more defensively he reacted in the form of denial of shortcomings, blaming others and various other forms of excuses. The investigators concluded that a new proposed system called the work-planning and review method, (WP&R), with its emphasis on mutual goal setting and problem solving, would help to alleviate this defensiveness. Earlier, McGregor (1957), in his analysis of the traditional performance techniques, had proposed a similar approach; placing greater responsibility on the subordinates for establishing performance goals and appraising progress toward them, thus stimulating the development of the subordinate. Burke and Wilcox (1969), in their review of the literature concerning appraisal interview techniques, found that success was related to a high level of subordinate participation, a helpful and constructive attitude on the part of the supervisor, solution of job problems and the mutual setting of specific goals to be aceived by the subordinate in the near future.

The above studies exemplify the attempts that have be

been made to soften or minimize potential conflict when rater and ratee face each other with the ratings facing both of them. The basic problem lies in the differing conceptions held by each of the current job performance by the ratee and more specifically, the differing conceptions of the importance of certain job factors, traits and behavior which contribute to job performance. These differences could very well become disproportionate due to the importance the ratee attaches to the ratings, upon which some very real decisions could be based.

Performance appraisals can be viewed, then, as subjective evaluations reflecting rater or supervisors' ideas of how certain behavior by the ratee contributes to those characteristics which are essential to good job performance and naturally are not a reflection of subordinates' ideas about that same behavior and how it contributes to job performance. Rarely has the subordinate's own conceptions been regarded as something worth exploring. Some essence of these conceptions could be obtained quite easily by instructing employees to rate themselves using the same rating forms used by their superviosrs. It would be expected that these self-ratings, when compared with supervisors' ratings, would clearly and directly demonstrate the discrepancies between employee and supervisor perceptions of job performance. The nature and degree of this discrepancy in general and the difference of such between low-rated and

high-rated employees is to be explored in this study.

II. EMPLOYER-EMPLOYEE DISCREPANCIES IN PERCEPTION

General Patterns

The discrepancy between subordinate and superior (or management) perceptions of job performance is an element in a general pattern of a well documented discrepancy in the perception of subordinate needs, wants and values. In a well known study by Kahn (1959), it was found that management consistently overestimated the importance workers attach to wages and generally underestimated the importance of sociopsychological needs. Gluskinos and Kestleman (1970) had factory workers, management personnel, and office employees list a number of job factors in order of importance to them. They also requested that management and union representatives rate the same factors as they believed the employees would. Value hierarchies of the factory and office workers were quite different beyond general agreement on the need for steady work. Both management and and union leaders held inaccurate perceptions of employees' wants by overestimating their needs for material rewards and underestimating the importance workers attach to good relations with peers and supervisors.

A study by Bescoe and Lawshe (1955) compared the perception of consideration and initiation of structure in foremen by superiors and subordinates. That superiors and

subordinates have very different conceptions about what type of behavior is representative of these two traits is evidenced by the findings that superior-perceived consideration and subordinate-perceived consideration correlated only .02. Likewise, the two groups' perception of initiation of structure correlated .04. An involved study by Jensen and Morris (1960) supports the belief that superiors do not perceive in their subordinates as a group those abilities which superiors value most in their contribution to overall job performance. Forty-two male janitors were ranked by their supervisors according to overall job performance. The janitors were then rated on the 120 items of the Wrigley scale. Items correlating positively with the original rankings were scored as +1, while those correlating negatively were scored as -1. Algebraic totals for each of the ten factors measured by the Wrigley scale were converted to percentages of the highest scores possible. On the basis of these percentage scores, two extreme groups were chosen. The individual ten factor scores for these two groups were then compared to each other and to the scores of the total group. Relative importance of the factors as perceived by the superiors was shown by the order of the size of the differences between the superior and inferior groups. A correlation of -.92 was found between mean factor scores of the total group and the order of importance of those factors as evidenced

by the ratings of superior and inferior groups. This indicates that the abilities valued most by the supervisors were least abundant in the workers.

Differences in perception of job performance: self-ratings

Comparatively little research has been performed on the differences in perception of job performance between superiors and subordinates in the industrial setting by comparing self-appraisals by employees with appraisals by their superiors. As stated earlier, such an approach would conceivably offer clues as to the exact nature and degree of the discrepancies between superior and subordinate perceptions of job performance. At least one investigator has stressed the importance of such research. Thornton (1968) states that self-appraisal systems are gaining more and more prominence and that

> the relationship between supervisory ratings and self-ratings has important implications for the usefulness of self-appraisal programs. If self-evaluations are to be effective in enlisting a person's cooperation and participation, it seems essential that his perceptions agree with those of his superior. This study was designed to identify those areas of executive performance where there might be significant amounts of disagreement. Knowledge of these problem areas would then be useful in conducting self-appraisal programs. (p.441)

It should be noted that Thornton was concerned with this problem at the executive level, but it is quite clear that such research has important implications for lower level employees also. In addition to the importance this type of research would have on the future use of self-evaluation,

which is an attempt to foster subordinate participation in the appraisal process, it would serve also as a starting point from which management can evaluate the effectiveness with which it communicates what type of behavior it believes best contributes to job effectiveness and, ultimately, organizational effectiveness. Maher and Piersol (1970) found that to the extent that an individual clearly perceives his individual job objectives, his satisfaction with the job will be high. The authors view this as a satisfaction of a higher order need which should be attended to now that increasing affluence is satisfying lower order needs. Self-ratings might give management the feedback it needs to further clarify job objectives.

Self-ratings have gained what little prominence they have not so much as an alternative to traditional rating systems, but rather as a research tool. Although self-ratings have been used in conjunction with supervisory ratings in a number of studies (Bassett and Meyer,1968), (Hall,1951), (Robinson,1970), there is a general tendency of employees to understandably modify their ratings downward somewhat if it is known that their supervisor is to see the ratings and that they may be used in a performance appraisal interview. As a result, this method yields modified ratings rather than true self-ratings. It becomes clear that if research into the discrepancy between supervisor and subordinate perceptions of job performance is to

be of any interest or value, then the opportunity ought to be presented to supervisors and employees to keep their ratings confidential, hopefully minimizing the need on the part of either group to modify their ratings in any way. Such an approach was chosen for this study.

Much of the research associated with self-ratings has used them in conjunction with supervisory ratings for use in a problem solving, goal-setting performance appraisal interview. Although, as stated above, such a method yields somewhat modified ratings. it is still possible to draw some conclusions about characteristics of self-ratings, and the discrepancy between self-ratings and supervisory rating results. One of these characteristics is, of course, the tendency of employees to overrate themselves. Meyer, Kay and French, in trying to explain the defensiveness in performance appraisal interviews, stated that this tendency on the part of employees might very well account for a good proportion of this defensiveness. The average employee's self estimate of performance before the interview placed him at the 77th percentile. Only two out of the 92 participants estimated their performance to be below the average point on the scale used. It is interesting to note that, after the interview, the same men, when asked how they thought their bosses had rated them, gave a figure that averaged at the 65th percentile. Parker, Taylor, Barrett and Martens (1959) also found that employees on the

whole rated themselves above their supervisors' ratings on all seven of their job performance factors and on the eighth over-all performance factor. Thornton, in his study, also found the same phenomenon among management personnel. Parker and his associates also found that estimates of supervisory ratings by employees on the average were consistently lower than self-ratings on the same factors, but still higher than supervisors' actual ratings. So it appears that, while employees realize that their self-ratings are higher than supervisors' ratings, their estimate of supervisors' ratings reveal that they do not realize the discrepancy is as great as it is.

On the basis of these studies, the following hypothesis was proposed: On the whole, employees will rate themselves higher than their estimate of supervisors' ratings, which , in turn, will be higher than the actual ratings by supervisors.

Group Differences

It is expected, of course, that there will be differences among employees and groups of employees in the degree of discrepancy between self-ratings and supervisors' rating scores. The subject of individual or group differences in this area has received little attention. One investigator, Musella (1969), found that with student teachers, close-minded subjects tended to rate themselves higher than open-minded subjects, suggesting at least one personality

variable at work. Although the area of personality variables promises to stimulate new interest and research, it was decided that the present study would compare groups of employees according to their standings on supervisors' ratings; that is, to compare low-rated employees with high-rated employees. Thornton took this approach in his study on management personnel. He found that, in general, low-rated employees rated themselves higher than their supervisors rated them to a greater degree than did high-rated employees. No similar approach, to the knowledge of the writer has ever been taken with non-management personnel. It was expected, however, that the same pattern would emerge. Thus, it was hypothesized; both high and low-rated employees will rate themselves higher than their estimates of their supervisors' ratings, which will be higher than the actual ratings by supervisors, but high-rated employees will show less discrepancy between their ratings (estimate) and supervisors' ratings than will low-rated employees.

Group Differences: Importance of Job Factors

Certain job factors and traits of employees contribute in varying degrees, of course, to an overall rating of general job performance. Factors on Format III such as Quantity of work or Quality of work would seem, at first glance, to contribute more to overall performance than would factors such as Cooperativeness or Conscientiousness.

Employees can be rated on a number of traits or factors on a graphic rating scale similar to Format III and the relative weight of each factor can be systematically determined by comparison with an overall estimate of general job performance. Parker and his associates used multiple regression to empirically determine the weights each group gave to each of the factors on Format III. A similar approach was selected for the present study. Using this method, it is possible to ascertain the relative weights given to certain selected factors by employees in their rating of themselves, their estimate of how their supervisors rate them, and the weight given those same factors by supervisors in their ratings of employees. Parker and his colleagues did just that in the study referred to above and discovered differences between supervisors and employees as a whole. One purpose of this study was to compare the relative weights given factors by low-rated and high-rated employees with those given by their supervisors and with each other.

One plausible reason why some employees are given high ratings by their supervisors might be that the relative importance attributed to factors in their contribution to overall performance given by these employees might be more similar to the importance their supervisors attach to these same factors. This line of thought leads naturally to a third hypothesis; the relative weights given

certain factors in their contribution to overall performance will differ or show more discrepancy between ratings by low-rated employees and supervisors' ratings than between ratings by high-rated employees and supervisors' ratings.

III. STATEMENT OF HYPOTHESES

From ratings gathered by the graphic rating device, Format III, by supervisors rating employees, by employees rating themselves, and by employees estimating what their ratings by their supervisors will be, the following hypotheses are proposed:

Hypothesis I

Employees rate themselves higher than their estimate of supervisors' ratings, which, in turn, will be higher than the actual ratings by supervisors.

Hypothesis II

Both high and low-rated employees will rate themselves higher than their estimate of supervisors' ratings which will be higher than the actual ratings, but high-rated employees will show less discrepancy between their estimate ratings and supervisors' ratings than will low-rated employees.

Hypothesis III

The relative weights given certain factors in their

contribution to overall performance will differ or will show more discrepancy between ratings by low-rated employees and supervisors' ratings than between ratings by high-rated employees and supervisors' ratings.

CHAPTER V

METHOD

I. THE RATING FORM

The basic rating form is identical to Format III. Format III, as described earlier, consists of eight rating scales, seven of which are designed to cover distinct areas of job performance and the eighth being an overall, general job performance rating. The rating form, without instruction pages, consisted of four 8½ by 11 inch pages, each page with two of the rating scales. Each scale was ten inches long, divided into five equidistant divisions, each with three further subdivisions.. The trait name for each scale appears above and to the left of the scale. The verbal anchoring appeared directly below the five major divisions. A copy of the rating form appears in the appendix.

II. SUBJECTS

Ratees consisted of 78 female production workers in eight different work groups at Tektronix, a large electronics manufacturing firm located in Beaverton, Oregon. Subjects engaged in essentially the same type of work involving the assembly of oscilloscopes. Raters consisted of

the eight immediate supervisors, all male, of each of the groups of employees mentioned. Company officials provided evidence that the groups did not essentially differ from one another in educational background, rate of pay and rate of production.

III. PROCEDURE

Each supervisor was given one rating form for each employee under him along with an instruction sheet, a copy of which may be found in the appendix. In short, supervisors were asked to rate each employee under his supervision using the given rating forms following the general outline included with the instructions. No strict time limit was given for completion of the ratings, although each supervisor was asked to return the completed ratings within a reasonable amount of time.

Each employee participating in the study was given two rating forms, each with its own instructions. The first set was used for the employees' self appraisals, while the second set was used for the employees' estimates of how they thought their supervisors had rated them.

Eight different half-hour sessions were held throughout one workday morning for the purpose of administering the two rating forms to the employees. Participants in the study were pre-assigned to a particular session by their supervisors and were told only to report to a

conference room near their work area at their assigned time. Employees in each session were seated around a large oval table and were given a brief oral presentation, a copy of which may be found in the appendix. Following the presentation, each participant was given a copy of the first set of ratings and asked to raise their hand upon completion, after which they were given the second set of ratings. The employees were not told the purpose of the study, nor were they aware of the instructions for the second set of ratings before they had completed the first set. No time limit was given, although most forms were completed well within 20 minutes. Participants in the study were told they could leave the room after completion of the second set of ratings and were asked to keep details of the study from their co-workers.

Scoring

The raw score on each rating scale was determined by measuring the distance in sub-division units from the left end of the scale to the point on the scale where the rater had placed a checkmark. All distances were rounded off to the nearest tenth of a unit.

CHAPTER VI

RESULTS

I. HYPOTHESIS I

Means and standard deviations of the scores of each of the eight scales for the three types of ratings were computed and are presented in Table I.

TABLE I

MEANS AND STANDARD DEVIATIONS OF SCALE SCORES OF THE THREE TYPES OF RATING USING FORMAT III (N=78)

Scale	Type of Rating					
	Self		Estimate		Supervisor	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1. Ability to work with others.	9.0	2.3	8.8	2.3	8.7	2.6
2. Amount of work done.	9.8	2.0	9.4	3.0	9.3	2.3
3. Quality of work done.	9.9	2.0	9.6	2.0	9.6	2.1
4. Leadership potential.	7.1	3.0	6.8	3.0	5.9	3.1
5. Ability to do complicated jobs.	8.9	2.3	8.4	2.2	7.9	2.0
6. Ability to work with minimum supervision.	11.6	2.1	10.7	2.4	9.5	3.0
7. Conscientiousness.	10.4	2.4	9.6	2.4	9.5	2.4
8. Overall performance.	10.1	2.0	9.9	1.4	9.6	2.2

As can be seen from the data, the scores for all scales except 3, Quality of work done, follow the pattern predicted in Hypothesis I; that is, the mean of the self

ratings was greater than the mean of the estimate ratings which were, in turn, greater than the mean of the supervisors' ratings. This general ordering of means was tested using Page's (Page, 1963) test for the ordering of means. The above mentioned pattern was found to be significant for scales 4 ($L=971$, $P<.05$), 5 ($L=983$, $P<.01$), 6 ($L=1003$, $P<.001$) and 7 ($L=977$, $P<.01$). The hypothesized ordering of all scores on scale 8, Overall Performance, was tested using Page's test and also found to be significant ($L=979$, $P<.01$).

II. HYPOTHESIS II

The division of the 78 ratees into high, low and middle thirds was based on the standard scores of the individual rating scores using the means and standard deviations of the scores of the eight supervisors. The means and standard deviations of the scores of each of the three ratings for the highest and lowest thirds are shown in Table II along with the mean differences between self-estimate rating and supervisors' ratings.

Hypothesis II predicted that the lower group would show more discrepancy between their estimate ratings and supervisors' ratings than would the higher group. Though the discrepancies for the two groups are in opposite directions (higher group underestimated, lower group overestimated), the difference in the absolute amount of discrepancy

TABLE II

RATINGS ON FACTOR 8 FOR UPPER AND LOWER THIRD EMPLOYEES

	Type of Rating						
	Self		Estimate		Supervisor		Mean Discrepancy Est.-Super.
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	
Upper third	9.75	2.2	9.53	2.4	11.89	1.3	-2.36
Lower third	10.03	1.7	10.02	1.6	7.40	1.2	+2.62

(.26) does not even approach significance. The lower group did not show significantly more discrepancy as had been predicted.

Hypothesis II also predicted that for both groups, self-ratings would be higher than supervisors' ratings, and that estimate ratings would be less than self ratings, but higher than the actual ratings by supervisors. This pattern was shown to be true only of the lower group ($L=342.5$, $P<.001$). The high group rated themselves significantly lower ($t=4.28$, $P<.001$) than their supervisors' ratings and estimated their ratings to be significantly lower ($t=4.37$, $P<.001$) than their supervisors' ratings.

That neither group of employees nor the employees as a whole demonstrated a moderate, much less a significant relationship between their self-ratings and supervisors' ratings is shown by the low correlation coefficients in Table III. Nor does there appear any significant increase

TABLE III

CORRELATIONS BETWEEN RATINGS ON FACTOR 8
FOR UPPER AND LOWER THIRD EMPLOYEES

	self.-super.	estimate-super.
Upper third	.12	.16
Lower third	.06	.16
All groups	.02	.22

in the relationship for either group in the shift from self to estimate ratings.

The results do not seem to indicate support for the hypothesis that high-rated employees show less discrepancy between their estimate ratings and ratings by their supervisors than do low-rated employees.

III. HYPOTHESIS III

Intercorrelations of rating scores for the eight factors used were computed for the self-ratings by high and low-rated employees as well as for the supervisors' ratings. Matrices of these obtained correlations are presented in the appendix. In order to determine the relative weights given by each of the three groups to the first seven factors in their contribution to Factor eight, Overall performance, the step-wise regression analysis method was chosen. This method first chooses the variable which contributes most to the variance of the dependent variable,

and then, in each additional step, adds another variable which next contributes the most variance and decreases the mean sum of squares. The final step of analysis includes all seven variables along with their relative beta-weights, each representing the approximate importance attached to that variable in its contribution to the overall performance rating by each group. Tables IV, V, and VI on the following pages are summaries of the step-wise analysis for each of the three groups. Included are the progressive steps in the multiple regression correlation coefficient, R and its square, R^2.

TABLE IV

SUMMARY OF STEPWISE REGRESSION ANALYSIS OF RATINGS BY HIGH-RATED EMPLOYEES

Variables	Beta-Coefficient	Multiple R	R^2
6. Ability to work with minimum supervision	.7772	.777	.604
2. Amount of work done	.4104		
6. Ability to work with minimum supervision	.6264	.866	.750
2. Amount of work done	.3722		
3. Quality of work done	.2253		
6. Ability to work with minimum supervision	.4955	.882	.778
2. Amount of work done	.4030		
3. Quality of work done	.2424		
6. Ability to work with minimum supervision	.5422		
7. Conscientiousness	-.0981	.884	.782
2. Amount of work done	.4079		
3. Quality of work done	.2329		
4. Leadership potential	.0774		
6. Ability to work with minimum supervision	.5482		
7. Conscientiousness	-.1383	.887	.786
2. Amount of work done	.3803		
3. Quality of work done	.2099		
4. Leadership potential	.0707		
5. Ability to do complicated jobs	.0702		
6. Ability to work with minimum supervision	.5281		
7. Conscientiousness	-.1331	.888	.788
1. Ability to work with others	-.0419		
2. Amount of work done	.3974		
3. Quality of work done	.2202		
4. Leadership potential	.0686		
5. Ability to do complicated jobs	.0736		
6. Ability to work with minimum supervision	.5248		
7. Conscientiousness	-.1354	.889	.790

TABLE V

SUMMARY OF STEPWISE REGRESSION ANALYSIS
OF RATINGS BY LOW-RATED EMPLOYEES

Variables	Beta-Coefficient	Multiple R	R^2
2. Amount of work done	.7685	.768	.591
1. Ability to work with others	.1857		
2. Amount of work done	.6841	.786	.618
1. Ability to work with others	.2647		
2. Amount of work done	.7721		
7. Conscientiousness	-.2095	.801	.642
1. Ability to work with others	.2973		
2. Amount of work done	.7812		
4. Leadership potential	.2309		
7. Conscientiousness	-.3807	.820	.672
1. Ability to work with others	.3347		
2. Amount of work done	.8433		
3. Quality of work done	-.1135		
4. Leadership potential	.2037		
7. Conscientiousness	-.3868	.824	.679
1. Ability to work with others	.4319		
2. Amount of work done	.9532		
3. Quality of work done	-.2477		
4. Leadership potential	.1026		
5. Ability to do complicated jobs	.2173		
7. Conscientiousness	-.4999	.837	.700
1. Ability to work with others	.4660		
2. Amount of work done	1.0126		
3. Quality of work done	-.3221		
4. Leadership potential	.1114		
5. Ability to do complicated jobs	-.2523		
6. Ability to work with minimum supervision	.1023		
7. Conscientiousness	-.5875	.841	.707

TABLE VI

SUMMARY OF STEPWISE REGRESSION ANALYSIS
OF RATINGS BY SUPERVISORS

Variables	Beta-Coefficient	Multiple R	R^2
7. Conscientiousness	.9071	.907	.823
6. Ability to work with minimum supervision	.3172		
7. Conscientiousness	.6593	.928	.862
2. Amount of work done	.2121		
6. Ability to work with minimum supervision	.2796		
7. Conscientiousness	.5288	.939	.881
2. Amount of work done	.1731		
3. Quality of work done	.1492		
6. Ability to work with minimum supervision	.2242		
7. Conscientiousness	.5080	.945	.892
2. Amount of work done	.1700		
3. Quality of work done	.1391		
5. Ability to do complicated jobs	.0635		
6. Ability to work with minimum supervision	.2005		
7. Conscientiousness	.4936	.946	.894
1. Ability to work with others	.0408		
2. Amount of work done	.1711		
3. Quality of work done	.1363		
5. Ability to do complicated jobs	.0600		
6. Ability to work with minimum supervision	.1822		
7. Conscientiousness	.4945	.946	.896
1. Ability to work with others	.0490		
2. Amount of work done	.1679		
3. Quality of work done	.1288		
4. Leadership potential	-.0344		
5. Ability to do complicated jobs	.0693		
6. Ability to work with minimum supervision	.1925		
7. Conscientiousness	.5046	.947	.896

To facilitate comparisons of the final beta-weights among the three groups, the final weights are presented in Table VII, along with the results of tests of significance of beta from zero, final multiple R and R^2s.

TABLE VII

BETA-WEIGHTS FOR EACH FACTOR WITH MULTIPLE CORRELATIONS (R) BETWEEN FACTORS 1 THROUGH 7 WITH FACTOR 8

Factor	Beta-weights		
	Low	High	Super-visor
1. Ability to work with others	.47*	-.04	.05
2. Amount of work done	1.01**	.40*	.17**
3. Quality of work done	-.32	.22	.13*
4. Leadership potential	.11	.07	-.03
5. Ability to do complicated jobs	.25	.07	.07
6. Ability to work with minimum supervision	.10	.52**	.19*
7. Conscientiousness	-.59*	-.14	.50**
Multiple R	.8408	.8886	.9466
R^2	.7070	.7896	.8961

* $P < .01$
** $P < .001$

As can be observed from Table VII, the high-rated employees seem to show more correspondence in their final

beta-weights to the supervisors than do the low-rated employees, with the notable exceptions of Factors 6 and 7, Ability to work with minimum supervision and Conscientiousness. The results for Factor 6 are deceiving, however. Among supervisors this was shown to be the most important factor after Conscientiousness, while with high-rated employees it was the single most important factor. Low-rated employees show it to be the least important factor. Both high and low-rated employees give Factor 7, Conscientiousness, the heaviest negative weight, while supervisors give it the highest positive weight.

Very high multiple Rs were found for all groups in the study, suggesting that the seven factors in question account for a large amount of the variance in the Overall performance ratings. Other factors accounted for 29% of the variance in ratings by low-rated employees, 21% in ratings by high-rated employees and only 10% in ratings by the supervisors. This suggests that there were other factors which contributed to job performance in the eyes of the employees of which the supervisors were unaware.

Few of the final beta-weights were found to be significantly different from zero; two among high-rated employees, three among low-rated employees, and four among supervisors. Individual beta-weights showed much less variance among supervisors, thus requiring lower beta values for significance.

As can be seen from Tables IV, V and VI, a limited number of factors can account for almost all the variance in Factor 8, Overall performance, contributed by all seven factors together. For high-rated employees, Factors 2,3 and 6, Amount of work done, Quality of work done and Ability to work with minimum supervision, together accounted for 77.8% of the variance, which was 98.6% of the variance accounted for by all seven variables. For low-rated employees, Factors 1,2,4 and 7, Ability to work with others, Amount of work done, Leadership potential and Conscientiousness accounted for about 67.5% of the variance or 95.1% of that accounted for by all variables. For the supervisors, Factors 2,3,6 and 7, Amount of work done, Quality of work done, Ability to work with minimum supervision and Conscientiousness accounted for 89% of the variance or 99.6% of the total variance accounted for by all seven factors. Conscientiousness alone accounted for 82% of the variance in Factor 8.

That the high-rated employees shared three of the four factors considered important by supervisors again shows the closer correspondence between conceptions of what contributes to job performance and that of their supervisors. Low-rated employees shared two of the four factors considered important by supervisors, but it must be pointed out that Factor 7, Conscientiousness, possesses a very strong negative weighting for low-rated employees,

but a very positive weighting for supervisors.

Factor 1--Ability to work with others--carries little or no weight with high-rated employees and supervisors but is a very definite contributor to job performance among low-rated employees. Apparently, low-rated employees overestimate the importance of their relationships with coworkers in its contribution to overall performance.

Factor 2--Amount of work done--is a consistent contributor to job performance across all groups. It carries the heaviest amount of weight of any of the factors covered among low-rated employees, and a fair amount of weight among high-rated employees and supervisors. All groups seem to recognize the importance of this factor in contributing to overall job performance.

Factor 3--Quality of work done--is given slight weight by high-rated employees and supervisors and a fair negative weight by low-rated employees, who seem to underestimate the contribution of this variable and indeed assign a negative weight to it, a result which is hard to interpret.

Factor 4--Leadership potential--carries little or no weight for high-rated employees and supervisors and only a slight weight for low-rated employees. None of the groups consider this factor to be of major importance.

Factor 5--Ability to do complicated jobs--carries little or no weight among high-rated employees and

supervisors, but a fair amount of weight among low-rated employees. Apparently, low-rated employees believe that if they have the ability to perform a job a bit more difficult than their present job, then it ought to contribute to present job performance.

Factor 6--Ability to work with minimum supervision--is given heavy weight by high-rated employees, moderate weight by supervisors and little weight by low-rated employees. High-rated employees seem to overemphasize and low-rated employees seem to underemphasize the contribution of this variable to overall performance in relation to that attributed by supervisors in their ratings.

Factor 7--Conscientiousness--is grossly underestimated by both groups of employees. Supervisors weighted this factor as the single most important in contributing to overall performance, while both groups of employees gave this factor negative weighting. This suggests that employees either have conceptions of what this factor means which are very different from that of their supervisors, or that they have perhaps similar conceptions, but disagree greatly on its importance.

It was decided to investigate the amount of discrepancy between mean scores between self-ratings and supervisors' ratings for the two groups of employees in relation to the ranked importance of the factors by supervisors as estimated from their relative beta-weights. The

Spearman rank correlations computed were .43 for low-rated employees and -.25 for high-rated employees. The differences in rank correlations was nonsignificant because of the small number of factors and only suggested that low-rated employees tend to show more discrepancy in those factors deemed important by supervisors.

CHAPTER VII

DISCUSSION

I. IMPLICATIONS OF THE STUDY

One finding of this study was that employees as a whole demonstrate a higher opinion of their overall job performance than do their supervisors. Employees as a whole also believe that their supervisors have a higher opinion of their job performance than they actually do. This result was not surprising in view of some of the results from past similar studies.

Probably one of the most important implications of this study is that there exist very definite group differences in how employees rate themselves and estimate their ratings by supervisors. In this particular study, groups of employees based on their ratings on an overall performance rating by their supervisors showed about the same amount of discrepancy between their ratings estimating their supervisors' ratings and the actual supervisors' ratings. But the most interesting finding was that the amount of discrepancy was in opposite directions, with low-rated employees overestimating themselves and high-rated employees underestimating themselves to a significant degree. If either group enjoyed a more accurate perception

of how their supervisors rated them, it certainly was not demonstrated in this study. Perhaps a tendency toward modesty in the high-rated group caused some downgrading of their self-ratings and estimate ratings. This tendency, along with a possible leniency effect on the part of supervisors which would increase their ratings might account for a good part of the discrepancy between the ratings of these two groups. Low-rated employees may have been demonstrating some sort of overcompensation for what they realized was their low standing among their fellow employees. A lack of leniency with this group of employees on the part of supervisors along with an overcompensation effect might account for a good deal of the discrepancy between these two groups.

Another explanation for the amount of discrepancy between estimated ratings and actual ratings for both groups might be that significant numbers of employees in both groups receive little or no feedback from their supervisors concerning their standing among fellow employees. This could be interpreted in a number of ways. One interpretation is that employees receive little or no feedback from their supervisors about their job performance at all. This would be a sorry state of affairs and one that is possible , but not probable. Another interpretation is that employees do receive feedback about job performance, but that it is the type of feedback that is without regard

to the performance of other workers. Perhaps an employee is given feedback about what she is doing well on the job but receives no information in the process about her relative standing in the group. If this interpretation were valid, the employee would have to use her own perceptions about her relative standing to rate herself or estimate her rating on the type of scale presented in this study. Conceivably, personality factors would be at work in this situation in conjunction with feedback from other sources than supervisors; possibly co-workers, pay-scales, etc.. This interpretation, in the opinion of the writer, deserves the most consideration. It may also be, however, that high-rated employees have very different ideas about the ability of their co-workers than do low-rated employees. High-rated employees, for example, may view other employees as generally below average workers, and themselves as slightly above average, but still above the majority of their fellow workers. Low-rated employees may have a higher opinion of their co-workers as being much above average, and may consider themselves below this level of their fellow workers, but still above average with "average" being very low within this particular group of workers.

The data obtained from this study does not fully support any of the above interpretations. Without further investigation well beyond the scope of this study, no explanation can be seriously attempted. The data do indicate

definite differences in how two groups of employees, classified according to how their supervisors rate them, estimate their supervisors' ratings; differences which data taken from employees as a whole do not suggest. From the results of this study, the notion that employees overestimate their performance appears to be a much too simple and broad generalization.

Though neither high nor low-rated employees show any more accuracy in estimating their ranking by supervisors on an overall performance scale, high-rated employeees do demonstrate more awareness of the importance attached to certain job factors by their supervisors. This lends some credence to the notion that high-rated employees are rated high partly because they share many of the same conceptions about the importance given those same job factors, The data do not indicate whether high-rated employees hold these same conceptions independent of any influence from their supervisors, or whether they are better able to discern their supervisors' ideals and bring their own job behavior in line with those ideals.

In summary, the data give definite evidence of the disparities between employees' concepts of their own job performance and that of their supervisors, along with evidence of a closer correspondence between high-rated employees and supervisors in the weight given to certain job factors in their contribution to overall performance than

between low-rated employees and supervisors. The data also indicate the inability of either group as a whole to estimate the ratings given by their supervisors accurately, a particularly surprising result when one considers that merit ratings by supervisors in this particular firm are used in a conscientious performance appraisal interview program.

The data also question the value of self-ratings for any other use than as a research tool and, in limited cases, as a basis for performance review discussions. A neglible correlation between ratings of themselves and by their supervisors was found among employees as a whole and in each of the groups of employees studied. This is strong evidence against replacing supervisors' ratings with self-ratings for purposes traditionally associated with performance ratings. Ratings by employees certainly demonstrate no more objectivity than do supervisors' ratings.

Self-ratings do have use, however, in research on employee perceptions about job performance. Information gathered from self-ratings can also give management feedback about employee-management discrepancies in perception. Whether any such discrepancies are of any importance depends upon management aims and objectives. Discrepancies found in the present study, for example, may or may not be of any consequence if management feels employees are

already working at their potential. On the other hand, management may feel that correcting some misconceptions on the part of employees may also clear up some misconceptions on the part of supervisors and nothing but good can result.

II. LIMITATIONS OF THE STUDY

The conclusions and implications of this study are, of course limited by a number of factors. The most basic limitation is probably the number and type of subjects. All subjects were female, and numbered but 78 in the total group. The division into three groups resulted in groups of only 26. This is short of any ideal number of subjects for a multiple regression analysis. The subjects were production workers, in contrast to most studies involving female workers, in which subjects are generally clerical workers. These results cannot be generalized to either the general population of production workers, nor to that of female workers. Also, at the time the study was conducted, the department in question was having an absenteeism problem and management admittedly was having a difficult time in communicating to the workers the importance of showing up for work consistently. This may have influenced ratings by supervisors in a indeterminable manner.

The rating scale itself demonstrated some limiting factors. A number of scales contributed little or nothing to the variance in Factor 8, Overall performance. Of

course, each group differed as to which factors they deemed important, and the inclusion of all seven factors, whether some of them contributed little or not, did make it possible for group differences to show themselves. Although no method was included in the framework of this study to check reliability of the scale, there was reason to believe that the scale was fairly reliable from evidence of earlier studies. However, in studies such as this, in which group differences are to be examined, precautions should be made to assure that any rating or evaluation system being used is reliable for all groups studied.

Also, this study ignored a potentially important group for comparative purposes, the middle third rated employees. As the study now stands, what conclusions there are that can be drawn are limited by the fact that they represent only high and low-rated employees.

II. RECOMMENDATIONS FOR FURTHER STUDY

This study leaves many questions unanswered. Perhaps with different and more specific approaches to some of the problems posed in this study, more may be learned about differences among employees and between employees and supervisors in rating job performance and in the importance of certain job factors. Different rating scales with different combinations of job factors might be used to discover if the same pattern found in this study exists

elsewhere. Larger numbers of subjects should be studied in a number of different circumstances. Perhaps the results of this study are unique with female production workers; only further research could give the answer.

One result of this study is that striking differences are suggested in the weights employees attach to job factors. In this study, employees were divided according to their ratings by supervisors. In future studies, employees could be differentiated by various methods. Employees working under a particular supervisor for a long period of time could be compared with those who have been under the same supervisor for a short period of time. It would have to be predicted that those in the former group would show more agreement with their supervisor's weighting of factors on similar scales. Groups of employees under differing types of supervision could be compared. Also, the effectiveness of a feedback-communication program could be evaluated in a closely controlled study.

Factors on which supervisors and employees as a whole or as a group taken on some basis from the total group differ considerably should be explored to determine how their respective perceptions differ. Conscientiousness, for example, in this study would be prime subject matter for this type of approach due to the abnormally large discrepancies demonstrated between employees and supervisors in their weighting of this factor.

Differences between employees who definitely underestimate themselves and those who overestimate themselves ought to be investigated to determine if personality factors are at play and to what extent they interact with other factors. One approach might be to have employees judge the performance of their fellow workers as a group. It may be that those rated high according to some independent measure and those rated low have very different ideas about how their coworkers rate as a whole, and where they themselves fit into the general distribution. Perhaps high-rated employees do not really underestimate themselves on an overall performance measure, but rate their fellow workers lower, while still considering themselves among the best workers of their group.

Finally, a similar study ought to be performed in a company that does not use performance review sessions with employees in a formal program to determine if a pattern similar to that found in this study results. Employees in such a situation would conceivably show even less correspondence between their self and estimate ratings, having received even less feedback or information about their job performance.

Any study performed in an industrial setting is limited by the inability to impose those experimental controls typical of laboratory studies. As a result, conclusions and implications are often not clearcut and definite.

Nonetheless, no other method exists to study people at work in a scientific manner. This particular study is only an introduction to one aspect of the promising study of group and individual differences among employees. Very little data exist in this area of research, and, as a result, the value of such research cannot be gauged by the present state of affairs, but rather what might follow in future studies.

CHAPTER VIII

SUMMARY

The purpose of this study was to gain some insight into the differing conceptions of job performance and job performance factors held by supervisors, employees in general, and of employees rated high and low on overall job performance by their supervisors. It was suggested that much of the potential conflict present when rater and ratee face each other with the ratings facing both of them is due to these differing conceptions of what good job performance is, and how certain factors and behavior contribute to that performance. These differences naturally become disproportionate due to the importance the ratee attaches to the ratings. Several studies were mentioned to exemplify the efforts to soften potential conflict between raters and ratees. It was noted that the discrepancy in the perceptions of job performance is an element in a general pattern of a well-documented discrepancy in the perception of subordinates' wants, needs and desires by superiors.

To assess employees' perceptions about their own job performance, self-ratings were taken, along with estimate-ratings of how employees thought their supervisors would rate them. A graphic rating scale was used, Format III, with seven job performance factors; Ability to work with

others, Amount of work done, Quality of work done, Leadership potential, Ability to do complicated jobs, Ability to work with minimum supervision, Conscientiousness, and an eighth scale, Overall performance. Serving as subjects were 78 female workers engaged in the assembly of electronic oscilloscopes along with their eight immediate supervisors, all male.

It was hypothesized, on the basis of earlier studies, that employees would rate themselves higher than their estimate ratings, which, in turn, would be higher than the actual ratings by supervisors. This general ordering was found to be significant for all factors except 1,2 and 3 at the .05 level and beyond. The ordering of scores on scale 8, Overall performance, was significant at the .01 level of significance.

It was expected that there would be differences among groups of employees in the degree of discrepancy between estimate-ratings and actual supervisors' ratings. Employees were divided into low and high groups according to the ratings given them by their supervisors, with both groups consisting of one-third of the total number of employees. It was hypothesized that low-rated employees would show more discrepancy in this case than would high-rated employees, with both groups overestimating their overall performance.

Though the discrepancies for the two groups were

found to be in opposite directions (higher group underestimating, lower group overestimating), the difference in the absolute amount of discrepancy does not even approach statistical significance. The lower group did not show significantly more discrepancy as had been predicted. Also, the high rated themselves and estimated their ratings to be significantly lower than actual supervisors' ratings.

The third hypothesis was based on the assumption that one reason high-rated employees are rated high is that the relative importance these employees attach to certain job factors is more similar to that attached by supervisors than those by low-rated employees. Multiple regression equations were computed for supervisors' ratings, high-self ratings and low-self ratings using the overall performance rating as the criterion and the other seven factors as predictor variables. It was found, as predicted, that high-rated employees did show closer correspondence in their weights to those given by supervisors than did low-rated employees.

Implications of the study were discussed, with possible explanations proposed for the unexpected results of the second hypothesis with no single explanation found to be satisfactory with the limited amount of data available. Recommendations for further study were suggested, particularly with different subjects, within a company with no appraisal interview, and with different methods of grouping

employees.

Some conclusions based on the study, limited by the nature and number of subjects and the type and number of factors used in the rating scale are as follows; a) employees, as a whole, rate themselves higher both on a self-rating and on an estimate of their supervisors' ratings than the supervisors actually do rate them, b) neither high nor low-rated employees have any better idea of how their supervisors rate them on overall performance; both groups show approximately the same amount of discrepancy between their estimate of their supervisors' ratings and the actual ratings, c) high-rated employees tend to underestimate themselves and low-rated employees tend to overestimate themselves on overall performance both to about the same degree, d) high-rated employees do seem to have a better idea of the job factors that are important in their contribution to overall performance than do low-rated employees; the factors high-rated employees judge to be important are generally the same factors their supervisors judge to be important with noted exceptions.

REFERENCES

Barrett, R.S., Taylor, E.K., Parker, J.W., and Martens, W.L. Rating scale content:I. Scale information and supervisory ratings. Personnel Psychology, 1958, 11,

Barrett, R.S. Performance rating. Chicago: Science Research Associates, 1966.

Bassett, G., and Meyer, H. Performance appraisal based on self-review. Personnel Psychology, 1968, 21, 421-430.

Bendig, A.W. The reliability of self-ratings as a function of the amount of verbal anchoring and the number of categories on the scale. Journal of Applied Psychology, 1953, 37, 38-41

Bendig, A.W. Reliability and the number of rating scale categories. Journal of Applied Psychology, 1954a, 38, 38-40

Bescoe, R.O., and Lawshe, C.H. Foremen leadership as perceived by superiors and subordinates. Personnel Psychology, 1959, 12, 573-583.

Bingham, W.V. Halo, invalid and valid. Journal of Applied Psychology, 1939, 23, 221-228

Bittner, R. Developing an industrial merit rating procedure. Personnel Psychology, 1948, 1, 403-32

Burke, R.J., and Wilcox, D.S. Characteristics of effective employmee performance review and development interviews. Personnel Psychology, 1969, 22 (3), 291-305

Ewart, E.S., Seashore, S.E., and Tiffin, J. A factor analysis of an industrial merit rating scale. Journal of Applied Psychology, 1941, 25, 481-86

Gluskinos, U.M., and Kestleman, B. Management and union leaders' perception of worker needs as compared with self-reported needs. Proceedings of the Annual Convention of the American Psychological Association, 1970, 5 (pt.2), 603-604

Grant, D.L. A factor-analysis of managers' ratings. Journal of Applied Psychology, 1955, 39, 283-286

Hall, W.B. Tools for improved personnel relations. New York: American Management Association, 1951, 39, 283-286

Jensen, M.B. and Morris, W.E. Supervisory ratings and attitudes. Journal of Applied Psychology, 1960, 44, 339-340

Kahn, R.L. Human relations on the shop floor. In E.M. Hugh-Jones (Ed.), Human relations and modern management. Chicago: Quadrangle Books, 1965

Maher, J.R. and Piersol, D.T. Perceived clarity of individual job objectives and of group mission as correlate factors of organizational morale. Journal of Communications, 1970, 20 (2), 125-133

McGregor, D. An uneasy look at performance appraisal. Harvard Business Review, 1965, 43, 123-129

Meyer, H.H., Kay, E.and French, J.R.P. Split roles in performance appraisal. Harvard Business Review, 1965, 43, 123-129

Musella, D. Perceptual-cognitive style as related to self-evaluation and supervisor ratings by student teachers. Journal of Experimental Education, 1969, 37 (3), 51-55

Page, E.B. Ordered hypotheses for multiple treatments: a significance test for linear ranks. Journal of the American Statistical Association, 1963, 56, 216-230

Parker, J.W., Taylor, E.K., Barrett, R.S. and Martens, L. Rating scale content: III. Relationships between supervisory and self-ratings. Personnel Psychology, 1959, 12, 49-63

Peters, D.L. and McCormick, E.J. Comparative reliability of numerically anchored scales versus job-task anchored rating scales. Journal of Applied Psychology, 1966, 50, 92-96

Roach, D.E. Factor analysis of rated supervisory behavior. Personnel Psychology, 1956, 9, 487-498

Robinson, D.B. Predicting police effectiveness from self reports of relative time spent in task performance. Personnel Psychology, 1970, 23 (3), 327-345

Seashore, S.E., Indik, B.P. and Georgopoulos, B.S. Relationships among criteria of job performance. Journal of Applied Psychology, 1960, 44, 195-202

Stockford, L. and Bissell, H.W. Factors involved in establishing a merit-rating scale. Personnel, 1949, Sept, 94-115

Symonds, P.M. On the loss of reliability in ratings due to coarseness of the scale. Journal of Experimental Psychology, 1924, 7, 456-461

Taylor, E.K. and Hastman, R. Relation of format and administration to the characteristics of graphic rating scales. Personnel Psychology, 1956, 9, 181-206

Thorndike, E.L. A constant error in psychological ratings. Journal of Applied Psychology, 1920, 4, 25-29

Thornton, G.C. The relationship between supervisory and self-appraisals of executive performance. Personnel Psychology, 1968, 21 (4), 441-455

Tiffin, J. and McCormick, E.J. Industrial Psychology. (5th ed.) Englewood Cliffs, N.J.: Prentice-Hall, 1965

APPENDIX A

MATRICES OF CORRELATION COEFFICIENTS

I. MATRIX FOR SUPERVISORS

	1	2	3	4	5	6	7	8
Factor 1	1.00	.35	.38	.51	.41	.53	.41	.47
Factor 2		1.00	.61	.42	.56	.65	.75	.79
Factor 3			1.00	.35	.57	.65	.62	.72
Factor 4				1.00	.58	.63	.59	.57
Factor 5					1.00	.68	.65	.70
Factor 6						1.00	.78	.83
Factor 7							1.00	.91
Factor 8								1.00

II. MATRIX FOR LOW-RATED EMPLOYEES

	1	2	3	4	5	6	7	8
Factor 1	1.00	.45	.48	.26	.04	.20	.57	.50
Factor 2		1.00	.58	.33	.12	.13	.59	.77
Factor 3			1.00	-.01	.26	.39	.30	.42
Factor 4				1.00	.51	.00	.64	.33
Factor 5					1.00	.11	.45	.11
Factor 6						1.00	.33	.04
Factor 7							1.00	.40
Factor 8								1.00

III. MATRIX FOR HIGH-RATED EMPLOYEES

	1.	2	3	4	5	6	7	8
Factor 1	1.00	.48	.35	.12	.39	.22	.29	.34
Factor 2		1.00	.38	.24	.61	.37	.56	.64
Factor 3			1.00	.36	.65	.64	.60	.69
Factor 4				1.00	.37	.34	.50	.38
Factor 5					1.00	.62	.59	.71
Factor 6						1.00	.70	.78
Factor 7							1.00	.65
Factor 8								1.00

APPENDIX B

INSTRUCTIONS

I. ORAL INSTRUCTIONS TO EMPLOYEES

Good Morning. My name is Harvey McGowan and I am a graduate student at Portland State University working on my master's thesis. As part of the work involved with my thesis, I am asking you this morning to complete two sets of ratings. The first set I will pass out to you shortly. It has instructions of its own on the cover sheet. Please note that the instructions refer to <u>two</u> sets of ratings attached. Actually there is only one. The second set of ratings referred to will be passed out once you have completed the first set of ratings. So, after you have completed the first set of ratings, raise your hand and I will give you the second set of ratings. Once you have completed those ratings, please bring them to the front of the room, after which you may leave. Please be assured that these ratings are for research purposes and will not be observed by anyone but myself. If there any questions now, or when you are working on the ratings, I shall be glad to answer them. You may begin as soon as you receive the first set of ratings. Thank you very much for your cooperation.

II. WRITTEN INSTRUCTIONS ACCOMPANYING FIRST SET OF RATINGS

Accompanying this instruction sheet are two sets of rating forms. You are asked to complete the first set and finish the set before going on to the second set of ratings.

The first set of ratings consists of seven different job factors or traits and an eighth overall performance rating. Simply place a checkmark at the point along the line which you think best describes your own performance. Here are some guidelines which should make your ratings more accurate and easier for you:

A) Try to consider each of the seven traits or factors separately; one at a time. Try to keep only one factor in mind when rating and try not to let ratings on any other factor influence you.

B) Do not hesitate to use the extremes on the scales, if they do, in fact, describe your performance as you see it. However, you should keep in mind that most employees probably are rated somewhere around the middle of the scales.

C) Do not labor over these ratings and spend too much time on them. Spend enough time on each scale to make what you think will be an appropriate judgment. Studies have shown that if too much time is spent on ratings, raters become confused and the ratings, then become confusing.

Remember that for this first set of ratings, you are rating yourself. Try to be as honest as possible, and do not read the instructions for the second set of ratings until you have completed the first set.

Thank you for your cooperation. If there are any questions now or while you are doing the ratings, please

raise your hand and someone will assist you. If you have no questions, please begin.

III. WRITTEN INSTRUCTIONS ACCOMPANYING SECOND SET OF RATINGS

As we all know, not all people see the same thing in the same way. You have just rated yourself on a number of traits and job factors as you see yourself. Your supervisor, though, may or may not have the same idea of how you rate on each of these factors. Would your supervisor rate you differently on this same rating form ?

For this set of ratings, you are asked to estimate how you think your supervisor would rate you. The directions are essentially the same as the instructions for the set you just finished, except, of course, that you trying to rate yourself as you think your supervisor would. So you should not worry about how you just rated yourself on the previous set of ratings, but concentrate on how you think your supervisor would rate you.

If you have no questions, go ahead and begin this set of ratings.

IV. INSTRUCTIONS TO SUPERVISORS

Instructions to supervisors were essentially the same as those accompanying the first set of ratings for employees, with the same guidelines and instructions; the only changes being those mentioning self-rating.